

II. PSYCHIATRIC DISORDER IN GENERAL POPULATIONS: PROBLEM OF THE UNTREATED "CASE"

Bruce P. Dohrenwend, Ph.D., F.A.P.H.A.

AT least 35 different investigators have attempted to count untreated as well as treated cases of psychiatric disorder in 44 community studies.¹ The majority of these have been published in the last 20 years and, reflecting expanding conceptions of what constitutes a "case," have reported rather high rates for the "true" prevalence of psychiatric disorder. The median rate, for example, in the 26 studies published in 1950 or later is 15.6 per cent, by contrast with only 2.1 per cent in the 16 studies published before 1950.²

Evidence for the validity of the measures of disorder used in these studies is sparse. The investigators have asked that their determinations be taken largely on the basis of the clinical training and experience of the psychiatrists who made the psychiatric assessments.³ This is true, by and large, even of the two most methodologically explicit and ambitious of these works—the Midtown Manhattan Study⁴ conducted in New York City, and the Stirling County Study in a maritime province of Canada.⁵

In both the Midtown Study and the Stirling County Study, the procedure for identifying "cases" was to have psychiatrists rate the written records of interviews that had been conducted by lay interviewers with probability samples of adults from the communities being studied. In the Stirling County Study, disorder was defined in terms of judged similarity to descriptions in the 1952 Diagnostic and Statistical Manual of the American Psychiatric Association.⁶ The

main measure was a psychiatric evaluation of "caseness" described as ". . . a rating of the probability that, at some time in his adult life, up to the time of the interview, the individual would qualify as a psychiatric case."⁷ The Stirling County researchers concluded ". . . that at least half of the adults in Stirling County are *currently* suffering from some psychiatric disorder defined in the APA Diagnostic and Statistical Manual."⁸

The main rating in the Midtown Study ranged respondents on a scale from "well" through five degrees of severity of symptomatology: "mild," "moderate," "marked," "severe," and "incapacitated."⁹ Almost a quarter (23.4%) of the Midtown respondents were classified in the last three categories—marked, severe, or incapacitated.¹⁰ These are referred to collectively as "impaired," and are the "cases" in the Midtown Study. Michael, one of the evaluating psychiatrists, put it this way:

"The individuals in the impaired category of mental health . . . are represented as being analogous to patients in psychiatric therapy. . . . When it is urged that the mental ratings "marked" and "severe" are comparable to the clinical conditions of patients in ambulatory treatment, and the rating "incapacitated" to the clinically hospitalized, the distinction is presented . . . as an attempt to anchor our conceptualizations in relation to known degrees of psychopathology."¹¹

In the Midtown Study, in which the impairment rating is a less inclusive definition of disorder than the Stirling County "caseness" rating, almost three-

quarters of the "cases" had never been in treatment with a psychiatrist or other professional psychotherapist.¹²

Problem

Thus, major conclusions from these studies are that substantial proportions of the general population are psychiatric cases and that the large majority of these cases have never been in treatment. The underlying and largely untested assumption¹³ is that the psychiatric conditions of typical untreated "cases" are similar to those of typical psychiatric patients—mainly outpatients—as these are encountered in clinical experience or described on the basis of such experience in the APA Manual.

In a previous paper, we reported our finding that large minorities of respondents sampled from community groups report as many symptoms of the kinds asked about in the Midtown and Stirling County studies as do psychiatric patients.¹⁴ The symptoms these respondents from nonpatient groups report, however, are in general of the less severe variety as judged by psychiatrists. Moreover, the same symptoms tended to be judged less serious by those in the community groups who reported them than by their patient counterparts. The purpose of the present paper is to carry further the investigation of whether typical untreated "cases" in the general population do in fact suffer from psychiatric conditions comparable to those of typical cases in psychiatric treatment.

In doing so, we shall assume that the Midtown clinicians, in referring their impairment ratings to outpatients, either did not mean or should not have meant Manhattan's largely upper- and upper-middle-class patients of private therapists. The reason is apparent in the Midtown researchers' own data; these show that almost half (48%) of the 40 current outpatients who fell into their sample were not placed in any of

the impaired categories on the scale by the Midtown psychiatrists.¹⁵ These outpatients were apparently mainly upper-class or upper-middle-class and probably treated in private offices, if the Midtown treatment census and the class distribution of the impaired among them are any indication.¹⁶ In terms of the Midtown Study's own psychiatric ratings, they are not severely ill enough to provide a good criterion group for impairment on the Midtown Study scale.

Method of Procedure

The data for the present inquiry are drawn from a methodological study of psychiatric disorder in contrasting class and ethnic groups in Washington Heights, which is located in the northern portion of Manhattan Island and contains about 270,000 people. In the larger study, different types of interviews and different types of measures of disorder are being compared on a controlled basis with around 500 persons consisting mainly of the following groups: 67 community leaders; 257 adult heads of families (both men and women, married and single) sampled on a probability basis from the general population of Washington Heights; 118 outpatients from various psychiatric clinics in or adjacent to Washington Heights; and 48 inpatients admitted to either of two mental hospitals.¹⁷ These represent completion rates of 79 per cent of the leaders, 66 per cent of the community sample respondents, and 76 per cent of the patients whom we set out to interview. All subjects were between 21 and 64 years of age, and a more complete description of them and of the over-all design of the study is set forth elsewhere.¹⁸

Our aim in selecting the community and patient respondents was to give equal representation to white Protestants of old American ancestry, Jews, Irish, Negroes, and Puerto Ricans.

Within the community sample, we attempted to balance educational levels within each ethnic group—with only partial success. Within the patient group, we wanted to secure examples of a wide variety of types of behavior problems. Our procedure was to have therapists or evaluators at the various facilities sort their patients on the basis of similarity to brief descriptions of the types of symptom behavior used by Star and others in studies of attitudes towards mental illness.¹⁹ We then selected patients to fill quotas for these descriptions. About half the clinic patients were on waiting lists or had just been admitted to a current course of treatment; the remaining half had been interviewed by lay interviewers an average of two years before, in one of our previous studies.²⁰ All of the hospital patients interviewed had been admitted within the week prior to the interview.

The community leaders were chosen on the basis of formal positions of influence and on the basis of reputation, and all but a few were Jewish, Irish, Negro, or Puerto Rican. Included were state assemblymen, city councilmen, municipal court justices, businessmen, school principals, clergymen, and individuals who proved influential in various neighborhood action groups formed around civil rights issues.

Of the 67 leaders, 14.9 per cent said at the time of the interview that they were or had been in treatment with members of the mental health professions. The corresponding figure for the 257 community sample respondents is 9.7 per cent.

These leader, patient, and community sample respondents were interviewed by 15 psychiatrists, all but one of whom had completed his residency training. The psychiatrists' assignments were randomized.

The respondents were also split at random between two different types of in-

terview instruments—one called the Structured Interview Schedule (SIS), the other the Psychiatric Status Schedule (PSS). Both questionnaires were designed to elicit evidence of psychiatric symptomatology and attendant impairment of functioning in work, marital and sexual relations, child rearing, housekeeping, friendship, and leisure activities. They differ sharply, however, in how they go about this coverage.

The Structured Interview Schedule is based on those used in the earlier Midtown Study, the Stirling County Study, and our own previous research in Washington Heights.²¹ It is a conventional type of survey questionnaire and relies heavily on items with fixed alternative response categories (e.g., "true-false") drawn, as in the earlier studies, principally from the World War II Neuropsychiatric Screening Adjunct²² and the MMPI.²³ The Structured Interview Schedule is, however, much expanded over the previous versions, especially in its coverage of impairment of functioning in role areas.

The Psychiatric Status Schedule, by contrast, relies mainly on open-ended questions, the probed responses to which are coded into fixed categories descriptive of pathology, on the basis of the clinical judgment of the interviewer. The PSS—developed by Spitzer, Endicott, and Cohen—has its origin in the traditional mental status interview that is routinely conducted as part of clinical examinations, but with the added attempt to standardize questioning and recording procedures.²⁴

Toward the end of the interview with either the SIS or the PSS, the psychiatrist interviewer made a series of global clinical assessments. These included a rating of "caseness" on the Stirling County scale and a rating of "impairment" on the rating scale developed in the Midtown Study. Additional ratings on these scales were made of the written records of the interviews with sub-

samples of the respondents, to investigate the effects of training by Stirling and Midtown staff in the rating systems, and the effects of knowledge of status of the respondents, including knowledge of whether they were patients.

Results

Consider first the ratings made by the psychiatrists on the basis of actually interviewing the respondents (Interviewer Ratings). The results on the Stirling "caseness" ratings are presented in Table 1. As would be expected, there is sharp contrast between the patient and nonpatient groups: Patients are rated far more likely to be cases than community sample respondents who, in turn, are rated more likely to be cases than are leaders. This is true regardless of whether the interview was conducted with the PSS or the SIS. Note, however, that there seems to be a stronger tendency for the psychiatrist-interviewers to see the community sample respondents as cases when they use the SIS—which is more like the original Midtown Study and Stirling County Study interview schedules. The results in Table 2 on the Midtown rating of impairment contain similarly sharp contrasts between the patient and nonpatient groups. Moreover, as would be expected, the inpatients are clearly rated as more impaired than the clinic outpatients. These contrasts also hold regardless of which interview schedule is used, although there is again a tendency for the psychiatrists to see the community sample respondents as sicker when using the SIS to conduct the interview.

Consider for a moment the small groups of patients and expatients among the community sample and leader respondents in Tables 1 and 2. These respondents tend to be rated much sicker than the nonpatients in these samples; we will look at them more closely a bit later. However, since small minorities of patients and expatients were included

in the Midtown Study and Stirling County Study samples for which rates were reported, we will combine leader and community sample respondents, who have patient histories, with nonpatients in these groups for most of the remaining analysis. The net effect of their inclusion, of course, is to slightly reduce the differences between the essentially nonpatient and patient groups. Nevertheless, this facilitates some useful comparisons of our results in Washington Heights with the results in Midtown Manhattan and Stirling County.

Thus, for example, when we ignore the distinctions in Table 2 between patients and nonpatients in the community sample, some arithmetic shows that between 31 per cent and 25 per cent, depending on whether we take the SIS or the PSS figure, were rated in the impaired categories as compared to the 23.4 per cent reported for Midtown. On the "caseness" rating, the Stirling County researchers saw considerably more illness in the rural maritime county they studied than our psychiatrists saw in the community sample from Washington Heights, New York City. The Stirling County figures were: A=30.5 per cent; B=24.6 per cent; C=26.2 per cent; and D=18.7 per cent.²⁵ To these ABCD ratings, the Leightons attached the following average "subjective probabilities" that the person rated would actually be a case if given a "full diagnostic investigation": A=.9; B=.7; C=.4; and D=.1. Thus, if we dichotomize the ratings into A and B as more likely to be cases versus C and D as less likely to be cases, "caseness" ratings for the Washington Heights community sample would indicate about 35 per cent for SIS respondents and 26 per cent for PSS respondents, as opposed to about 55 per cent for Stirling County.²⁶

There is an interesting question here as to why the Washington Heights community respondents appear so much

Table 1—Interviewer "caseness" rating according to status of respondent and whether Structured Interview Schedule (SIS) or Psychiatric Status Schedule (PSS) was used: leaders and community sample respondents divided into those with and without a treatment history (in per cent)

	Status of respondent											
	Leader				Community sample						Hospital inpatients	
	History		No history		History		No history		Clinic			
	SIS	PSS	SIS	PSS	SIS	PSS	SIS	PSS	SIS	PSS		
"Caseness" rating												
A=Almost certainly psychiatric case	50.0	25.0	5.7	9.1	60.0	46.7	19.3	13.6	76.3	77.6	100.0	95.7
B=Probably a psychiatric case (too vague to rate "A")	0.0	75.0	2.9	4.5	10.0	6.7	12.3	9.3	8.5	6.9	0.0	4.3
C=Possibly a psychiatric case (borderline)	16.7	0.0	31.4	22.7	10.0	33.3	34.2	39.0	13.6	10.3	0.0	0.0
D=No evidence of symptoms. Some evidence of good health and adjustment	33.3	0.0	60.0	63.6	20.0	13.3	34.2	38.1	1.7	5.2	0.0	0.0
Total per cent	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Total respondents	(6)	(4)	(35)	(22)	(10)	(15)	(114)	(118)	(59)*	(58)	(25)	(23)

* Excludes one respondent for whom no rating was made.

Notes: With ratings dichotomized into A+B vs. C+D, chi-square tests show that none of the SIS vs PSS differences in proportions likely to be "cases" is statistically significant at the 0.10 level or better, though there is a trend ($p < 0.20$) for more SIS community sample respondents to be rated in the A and B categories. By contrast, with SIS and PSS respondents combined, $p < 0.01$ that the over-all differences among the leaders, community sample, clinic, and inpatient respondents could have occurred by chance. More specifically, the leaders and community sample respondents both have significantly fewer As and significantly more Bs than the clinic and hospital patients. The leaders have significantly more Ds than the community sample respondents. And the hospital patients have significantly more As than the clinic patients. Each of these specific differences is significant at the 0.05 level or better.

Table 2—Interviewer impairment rating according to status of respondent and whether Structured Interview Schedule (SIS) or Psychiatric Status Schedule (PSS) was used: leaders and community sample respondents divided into those with and without a treatment history (in per cent)

Impairment rating	Status of respondent											
	Leader				Community sample							
	History		No history		History				No history			
	SIS	PSS	SIS	PSS	SIS	PSS	SIS	PSS	SIS	PSS	SIS	PSS
Unimpaired:												
Well	16.7	0.0	48.6	50.0	10.0	13.3	29.8	23.7	0.0	3.4	0.0	0.0
Mild	16.7	50.0	34.3	22.7	20.0	26.7	21.1	38.1	5.1	6.9	0.0	0.0
Moderate	33.3	25.0	11.4	18.2	10.0	6.7	21.1	16.9	10.2	8.6	0.0	0.0
Subtotal	66.7	75.0	94.3	90.9	40.0	46.7	72.0	78.7	15.3	18.9	0.0	0.0
Impaired:												
Marked	33.3	25.0	2.9	4.5	30.0	26.7	21.1	13.6	18.6	22.4	0.0	4.3
Severe	0.0	0.0	2.9	4.5	20.0	26.7	4.4	5.1	47.5	25.9	16.0	8.7
Nearly incapacitated	0.0	0.0	0.0	0.0	10.0	0.0	2.6	2.5	16.9	27.6	40.0	56.5
Incapacitated	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	5.2	44.0	30.4
Subtotal	33.3	25.0	5.8	9.0	60.0	53.4	28.1	21.2	84.7	81.1	100.0	99.9
Total per cent	100.0	100.0	100.1	99.9	100.0	100.1	100.0	99.9	100.0	100.0	100.0	99.9
Total respondents	(6)	(4)	(35)	(22)	(10)	(15)	(114)	(118)	(59)*	(58)	(25)	(23)

* Excludes one respondent for whom no rating was made.
 Note: With the ratings dichotomized into "impaired" and "unimpaired," chi-square tests show that none of the PSS vs. SIS differences in proportions impaired is statistically significant at the 0.10 level or better, though there is a trend ($p < 0.30$) for more SIS community sample respondents to be rated in the impaired categories; by contrast, with SIS and PSS respondents combined, $p < 0.01$ that the over-all differences in proportions impaired among leaders, community sample, clinic, and hospital patients could have occurred by chance. More specifically, the leaders have a significantly smaller proportion impaired than the community sample ($p < 0.05$) which, in turn, has a significantly smaller proportion impaired than either the outpatients ($p < 0.01$) or inpatients ($p < 0.01$). Moreover, the inpatients have a significantly ($p < 0.01$) higher proportion in the incapacitated categories than the outpatients.

closer to the Midtown sample respondents on the "impairment" rating than to the Stirling County sample on the "caseness" rating.²⁷ There is no apparent substantive explanation for this latter result in terms of differences among the populations studied. As we show elsewhere, the difference seems to be due in part at least to the fact that the ratings reported above for our Washington Heights research were made on the basis of a face-to-face interview rather than on the basis of a written record of the interview.²⁸

With regard to our problem of whether typical "cases" in the community suffer from psychiatric conditions comparable to those of psychiatric patients, recall that respondents rated in any one of the "impaired" categories were thought, by the Midtown Study psychiatrists, to show disorder similar to the clinical conditions seen in psychiatric patients. Those rated "marked" and "severe" were held to be comparable to patients in ambulatory treatment; those rated in either the "nearly incapacitated" or "incapacitated" categories, to the clinically hospitalized.

Table 2 shows that the Washington Heights results are in accord with what the Midtown clinicians would have expected for the large majority of the hospital patients: These were rated "nearly incapacitated" or "incapacitated" with either the SIS or the PSS. It is startling to find, however, that, contrary to these same expectations, the average impaired clinic patient is quite different from the average impaired respondent in the community sample.

These contrasts between the impaired respondents in the community sample and the impaired respondents in the clinic group are somewhat more pronounced when the SIS was used than when the PSS was used. They occur strongly, however, with both instruments. We can summarize them more simply, therefore, by combining impaired SIS and PSS respondents.

This is done in Table 3 which shows that, typically, an impaired community sample member in our Washington Heights study was placed in the "marked" category, as was originally the case in the Midtown Study as well.²⁹ By contrast with these "impaired" respondents in the Midtown and Washington Heights community samples, the large majority of the impaired outpatients in the Washington Heights study were placed in the "severe" and "incapacitated" categories.

When we look at impaired nonpatients and impaired patients from the community sample separately, we again find a difference in the severity of impairment. Some further arithmetic with the results in Table 2 will show that only 30 per cent of the former were rated as high as "severe" by contrast with fully 50 per cent of the latter. While this difference is not statistically significant ($p > 0.10$) with an n of only 14 impaired community sample respondents with patient histories, it is large and it is consistent for both SIS and PSS respondents.

Only among the leaders, with few respondents in any of the impaired categories, is there no evidence of less severe impairment among the untreated "cases" than among the treated "cases" on the Midtown scale. Possibly this is because the leaders with treatment histories are mainly atypical patients of therapists with private office practices—something the leaders can better afford than most of the other respondents. This would be consistent with the fact that the large majority of the mainly upper- and upper-middle-class current patients in the Midtown Study sample were either rated unimpaired or, if impaired, were placed in the "marked" rather than the "severe" or "incapacitated" categories.³⁰

Unlike the Midtown Study impairment rating, which assesses an aspect of the severity of disorder, the Stirling County Study's "caseness" rating carries no clear implication of severity. We

will, therefore, investigate the relation of the Stirling rating to the Midtown rating to test whether, by the Stirling rating, a typical "case" in the community resembles a typical case in the clinic or mental hospital. We would expect, for example, that if average community sample respondents rated "A"—that is, "almost certainly a psychiatric case"—suffer from disorders comparable to those of average clinic patients rated "A," then the two should have similar levels of impairment on the Midtown rating.

Table 4 shows, however, that this is not so. Proportionally, almost twice as many of the outpatient As are rated "severe" or higher than of the community sample As. And, as would be expected, the contrast is even sharper when the comparison is with inpatients. Although it is not shown in Table 4, the differences hold for both PSS and SIS respondents—though they are stronger for the latter.

It would seem then that, by the Stirling "caseness" rating as well as by the Midtown "impairment" rating, the

typical "case" in the community is quite different from the typical case in the clinic or mental hospital. On the basis of the results we have presented so far, however, there are two major problems with this interpretation. First, the interviewing psychiatrists in our Washington Heights study usually guessed correctly or learned outright that they were interviewing psychiatric patients when this was the case. Is it possible, then, that knowledge of patient status produced a kind of negative halo effect, leading the psychiatrists to rate their respondents more severely ill than they would have done if they had not realized they were patients?

Second, these psychiatrist-interviewers were trained by psychiatrists on our staff in the Midtown Study and Stirling County Study rating procedures on the basis of published accounts.³¹ Would there have been less contrast between "cases" in the patient and nonpatient groups if psychiatrists from the Midtown and Stirling County studies had made the clinical ratings?

To investigate these possibilities, we

Table 3—Interviewer rating of degree of impairment among impaired respondents from the leader, community sample, and patient groups: SIS and PSS combined (in per cent)

Degree of impairment	Status of respondents rated impaired			
	Leader	Community sample	Clinic	Hospital inpatients
Marked	71.4	66.2	24.7	2.1
Severe	28.6	23.9	44.3	12.5
Nearly incapacitated	0.0	9.9	26.8	47.9
Incapacitated	0.0	0.0	4.1	37.5
Total per cent	100.0	100.0	99.9	100.0
Total respondents rated "impaired"	(7)	(71)	(97)	(48)

NOTE: The important comparison in this table is between the community sample and clinic outpatients. With "nearly incapacitated" and "incapacitated" combined, chi-square tests show $p < 0.01$ that the over-all differences between respondents rated impaired in these two groups could have occurred by chance; moreover, the proportion of impaired clinic rated "severe" or higher is significantly larger ($p < 0.01$) than the proportion of impaired community sample respondents rated "severe" or higher.

Table 4—Interviewer rating of impairment for respondents rated “A”—almost certainly a psychiatric case—from the leader, community sample, and patient groups: SIS and PSS combined (in per cent)

Impairment rating	Status of respondents rated “A”			
	Leader	Community sample	Clinic	Hospital inpatients
Unimpaired:				
Well	0.0	0.0	0.0	0.0
Mild	0.0	0.0	0.0	0.0
Moderate	37.5	19.6	2.2	0.0
Subtotal	37.5	19.6	2.2	0.0
Impaired:				
Marked	37.5	39.2	21.1	0.0
Severe	25.0	27.5	43.3	12.8
Nearly incapacitated	0.0	13.7	28.9	48.9
Incapacitated	0.0	0.0	4.4	38.3
Subtotal	62.5	80.4	97.7	100.0
Total per cent	100.0	100.0	99.9	100.0
Total respondents rated “A”	(8)	(51)	(90)	(47)

NOTE: The important comparison in this table is between the community sample and clinic. With “well,” “mild,” and “moderate” combined, and with “nearly incapacitated” and “incapacitated” combined, chi-square tests show $p < 0.01$ that the over-all differences between respondents rated “A” in these two groups could have occurred by chance. Moreover, the proportion of “A” clinic rated “severe” or higher is significantly larger ($p < 0.01$) than the proportion of “A” community sample respondents rated “severe” or higher.

shall turn to data from a second set of ratings (Reviewer Ratings) that were made on a subsample of the written records of the original interviews. It will be recalled that psychiatrists on the Midtown Study and the Stirling County Study made their ratings from written records of interviews conducted by lay interviewers. It was thus possible for us to have small groups of our own psychiatrists trained by psychiatrists from the Midtown and Stirling County studies on data and original ratings from these studies. This special training proceeded to the point where several of our psychiatrists showed satisfactory reliability with the original ratings from the Midtown and Stirling County studies.⁸²

It was also possible to edit the written records of the Washington Heights interviews to remove clues to patient status (e.g., drug history, reference to treatment, and the like) and other background characteristics. This was done on a subsample of something over a fifth of the respondents interviewed in our Washington Heights Study, to prepare the protocols of the interview for review.

In this review, 15 community sample respondents were rated “A” on the Stirling County scale, in a consensus of two of our reviewing psychiatrists who had been trained to be reliable with Stirling County Study psychiatrists.⁸³ Of those, only a third were rated as high as “severe” by a third psychiatrist who con-

tributed an independent Midtown Study impairment rating after showing very high reliability with the Midtown Study psychiatrists.³⁴ By contrast, of the 27 clinic patients rated "A" on the Stirling scale, 63 per cent were rated "severe" or higher on the Midtown scale in these reviews. Similarly, among the 14 community sample respondents rated in one of the impaired categories in this review, only 36 per cent were rated as high as "severe." Of the 27 clinic patients in the impaired categories in this review, 63 per cent were rated "severe" or higher. This difference and the previous one are significant at the 0.10 level with a one-tailed test.³⁵ Both differences are also large and wholly consistent in direction with the findings from the interviewer ratings.

In these review ratings, it was possible to add an additional rating from the Stirling County Study by a reviewer trained in that study's rating system.³⁶ This additional rating was also an "impairment" rating which differs somewhat from the Midtown impairment rating but, on the basis of our results, also has much in common with it. The Stirling County impairment rating was made in five categories: "none," "minimal," "mild," "moderate," and "severe." The first two categories, "none" and "minimal," were held to constitute "negligible impairment," while the last three were held to constitute "significant impairment."³⁷ Of the 15 community sample respondents in the review who were rated "A" on "caseness," only 20 per cent were rated as high as "moderate" on the Stirling impairment scale; by contrast, 59 per cent of the 27 "A" cases among the clinic patients were rated moderate or higher—a difference that is significant at the .01 level with a one-tailed test.

It might be argued, of course, that Washington Heights is not Midtown Manhattan, much less Stirling County, and that the results might be much dif-

ferent in those places so far as treated and untreated "cases" are concerned. Consider, however, that, for this to be so, either the nonpatient groups in Midtown and Stirling County would have to be much sicker than in Washington Heights, and/or the clinic and hospital patients much healthier. Neither of these possibilities seems very reasonable.

Still, it might be argued, the data collection procedures used in Washington Heights were not the same as those in either the Midtown Study or the Stirling County Study. Consider, however, that there is probably more difference between the PSS interview and the SIS interview in the Washington Heights study than between the SIS and the types of interviews used in the Midtown and Stirling County studies. Yet the differences between treated and untreated "cases" in the Washington Heights study hold for the PSS as well as the SIS respondents.

It would seem, then, that neither a negative halo effect resulting from knowledge of patient status, nor differences in comprehension of the rating systems from those held by the Midtown Study and the Stirling County Study researchers, could have accounted for the results obtained by our psychiatrist-interviewers. Nor can our results be easily explained away by differences in research setting or by differences in interview instruments. Rather, our research in Washington Heights strongly suggests that typical "cases" in the community, according to the Midtown Study and the Stirling County Study psychiatric ratings, are simply not the same as typical cases in either outpatient clinics or mental hospitals.

Implications

The most immediate implications of our findings center on what, in one sense, can be thought of as the tech-

nical problem of establishing cutting points on measures aimed at identifying psychiatric "cases" in epidemiological research. This problem certainly bothered the Stirling County Study and Midtown Study researchers, as both groups refer several times to the ambiguities involved in developing cutting points on the basis of untested clinical opinion.³⁸ On the basis of our research in Washington Heights, we have some suggestions to make on this matter of cutting points for investigators who are using or may plan to use the Stirling County Study or Midtown Study rating procedures.

If resemblance to typical clinic outpatients seems a reasonable criterion of minimal severity of psychopathology to count as a "case," the cutting point on the Midtown impairment scale should be raised to a point nearer "severe" than to "marked." If such a step were taken, the proportion of cases reported with current cutting points on the Midtown scale would probably be more than halved—if our research in Washington Heights in comparison with the results in Midtown are an indication. The rates according to the Stirling County "caseness" rating, if similarly modified in terms of that study's impairment rating, would be reduced much more than those rated according to the Midtown rating. The reason, of course, is that the Stirling "caseness" rating is the more inclusive of the two.

It is tempting to go from this discussion of the problem of establishing cutting points on these ratings on some reasonable empirical basis, to a discussion of the practical implications of the Washington Heights findings. Consider, however, that even with the lowered rates that would be "produced" by our suggested modification of cutting points on these scales, the ratio of untreated to treated cases in the general population would still be very high—staggering, in fact, given the available manpower for outpatient and inpatient serv-

ices. Thus, our recommendation would hardly provide either temporary solace to those charged with delivering services, or a precise guide for long-range public health planning in the field of mental health.

More important in the long run than the technical suggestion set forth above, let us suggest, will be intensive investigation of the bases for the differences between typical treated and typical untreated "cases" on these scales, and the nature of the overlap between them. For example: Are untreated "cases" with "marked" impairment usually milder varieties of the same general types of disorder as treated cases rated "severe?" Or are they rather earlier stages in the development of more severe types of disorder? Or are there indeed qualitative differences involved in typical untreated "cases" as opposed to typical treated cases?

When we have answers to such questions, there will probably no longer be a need for global measures such as the Stirling County "caseness" rating and the Midtown Study impairment rating. Rather, we will have much of the knowledge needed to develop more precise and objective measures that should prove more useful in our attempts to advance both theory and practice in dealing with problems of psychiatric disorder.

ACKNOWLEDGMENTS—The author would like to take this opportunity to thank the psychiatrists who did the interviewing on this study: Drs. Scott Angus, Edgardo Bianchi, Cor de Hart, Guy Edwards, Gladys Egri, Mearl Fenwick, Ursula Heyum, Elizabeth Jenks, Alan Levine, Isaac Lubchansky, Karo Manaser, Miriam Mathe, Fred Mendelsohn, Daniel O'Connell, and Sheldon Zimberg.

He would also like to express appreciation to Dr. Stanley Michael for his efficient and illuminating training of a group of our psychiatrists in the Midtown rating system, and Dr. Alexander H. Leighton for his generous provision of materials and staff for training in the Stirling County ratings. In this latter connection, Dr. Morton Beiser and, particularly, Dr. Daniel O'Connell—who was responsible for most of the actual training—were of immense

help. The psychiatrists who undertook the special training, especially Dr. Gladys Egri and Dr. Yorihiro Kumasaka, along with Dr. O'Connell, have made a major contribution to the study.

Finally, the author is grateful to Dr. Edwin Chin-Shong and Dr. Egri for their analyses of the data on the reliability of the ratings which were drawn on for this paper, and to Mrs. Karen Goodman for her able statistical assistance.

REFERENCES

1. Dohrenwend, Bruce P., and Dohrenwend, Barbara Snell. *Social Status and Psychological Disorder: A Causal Inquiry*. New York: Wiley, 1969, Chapter 2, pp. 9-31.
2. *Ibid.*
3. *Ibid.*, Chapter 7, pp. 95-109.
4. Srole, Leo; Langner, Thomas S.; Michael, Stanley T.; Opler, Marvin K.; and Rennie, Thomas A. C. *Mental Health in the Metropolis*. New York: McGraw-Hill, 1962.
5. Leighton, Dorothea C.; Harding, John S.; Macklin, David B.; Macmillan, Allister M.; and Leighton, Alexander H. *The Character of Danger*. New York: Basic Books, 1963.
6. American Psychiatric Association Committee on Nomenclature and Statistics. *Diagnostic and Statistical Manual: Mental Disorders*. Washington, D. C.: The Association, 1952.
7. Leighton, Dorothea C., et al. *Op. cit.*, p. 356.
8. *Ibid.*
9. Srole, Leo, et al. *Op. cit.*, p. 399.
10. *Ibid.*, p. 138.
11. *Ibid.*, p. 333.
12. *Ibid.*, p. 147.
13. Dohrenwend, Bruce P., and Dohrenwend, Barbara Snell. *Op. cit.*, Chapter 7, pp. 95-109.
14. Dohrenwend, Bruce P., and Crandell, DeWitt L. *Psychiatric Symptoms in Community, Clinic, and Mental Hospital Groups*. *Am. J. Psychiat.* 126:1611-1621 (May), 1970.
15. Srole, Leo, et al. *Op. cit.*, p. 147.
16. *Ibid.*, pp. 241, 246.
17. A group of 24 prisoners convicted of crimes in New York City were also interviewed as part of the study but are omitted from the analyses reported in this paper. The reason is that the prisoners are in an ambiguous status with regard to the Midtown Study definition of a case in terms of similarity to psychiatric patients.
18. Dohrenwend, Bruce P.; Chin-Shong, Edwin T.; Egri, Gladys; Mendelsohn, Frederick S.; and Stokes, Janet. "Measures of Psychiatric Disorder in Contrasting Class and Ethnic Groups: A Preliminary Report of On-Going Research." In: *Proceedings of the Aberdeen Symposium on Psychiatric Epidemiology*. Hare, Edward H., and Wing, John K. (eds.). London: Oxford University Press. (In press.)
19. Star, Shirley A. *The Public's Ideas About Mental Illness*. National Opinion Research Center, University of Chicago, 1955, mimeographed. For a published description of the types of symptom behavior see, for example, Dohrenwend, Bruce P., and Chin-Shong, Edwin T. *Social Status and Attitudes Toward Psychological Disorder: The Problem of Tolerance of Deviance*. *Am. Sociol. Rev.* 32:417-433 (June), 1967.
20. See, for example, Dohrenwend, Bruce P. *Social Status and Psychological Disorder: An Issue of Substance and an Issue of Method*. *Ibid.* 31:14-34 (Feb.), 1966.
21. *Ibid.*
22. Star, Shirley A. "The Screening of Psychoneurotics in the Army: Technical Development of Tests." In: *Measurement and Prediction*. Stouffer, Samuel A., et al. (eds.). Princeton, N. J.: Princeton University Press, 1950, pp. 486-547.
23. Dahlstrom, W. C., and Welsh, G. S. *MMPI Handbook*. Minneapolis: University of Minnesota Press, 1960.
24. Spitzer, Robert L.; Endicott, Jean; and Fleiss, Joseph L. *Instruments and Recording Forms for Evaluating Psychiatric Status and History: Rationale, Method of Development and Description*. *Comprehensive Psychiat.* 8:321-343 (Oct.), 1967.
25. Leighton, Dorothea C., et al. *Op. cit.*, p. 121.
26. *Ibid.*
27. The Stirling County Study investigators attempted to deal with past as well as current "caseness," whereas we attempted to deal only with current "caseness." However, the Stirling County researchers estimated that 90 per cent of their "cases" were current (*Ibid.*, p. 356). Thus their inclusion of past cases can hardly account for the difference.
28. Dohrenwend, Bruce P., et al. *Op. cit.*
29. Srole, Leo, et al. *Op. cit.*, p. 138.
30. *Ibid.*, p. 147.
31. The large majority of the interviews in the Washington Heights study were tape-recorded. As a check on the quality of the training on the basis of published accounts, one of the psychiatrists on our staff who was responsible for it, Dr.

Gladys Egri, made independent Midtown impairment ratings and Stirling County "caseness" ratings on a subsample of 69 of the tapes. The majority of these concentrated on the first few interviews each psychiatrist did, so that unreliable trends could be corrected at the outset. In brief summary: For the Stirling "caseness" rating, there was complete agreement 69.6 per cent of the time and one step disagreement 21.7 per cent of the time by contrast with 65 per cent and 31 per cent, respectively, reported by the Stirling County researchers. However, the agreement in Washington Heights for patients and leaders was greater than for the 35 community sample respondents in the spot check, who are really more comparable to respondents in the Midtown and Stirling samples. For the community sample, complete agreement was 60.0 per cent and one step disagreement was 28.6 per cent in the Washington Heights Study. On the Midtown impairment rating, complete agreement in the Washington Heights check was achieved 52.2 per cent of the time, and one-step disagreements 37.7 per cent of the time, by contrast with respective figures of 47.2 per cent and 44.7 per cent reported in the Midtown Study. Again, however, the reliability was less good for the Washington Heights community sample, where complete agreement was 42.9 per cent and one-step disagreement 43.8 per cent. Nevertheless, the degree of reliability seems comparable to that achieved in the Midtown and Stirling County studies, especially when it is realized that the Washington Heights situations of the original rating (actual interview) and review rating (from tape recordings) are not as comparable as when both sets of ratings are made from the same written record as in the Midtown and Stirling investigations.

32. Two psychiatrists from the Midtown Study, Dr. Stanley T. Michael and Dr. Price Kirkpatrick, did the psychiatric ratings on that study. Dr. Michael conducted a series of training sessions for three of our psychiatrists during which they rated a sample of 43 interview protocols that had been rated by Michael and

Kirkpatrick in the Midtown Study. Two of our three psychiatrists proved quite reliable with the Midtown psychiatrists, and one of them, Dr. Yorihiro Kumasaka, actually proved more reliable with Michael than Michael was with Kirkpatrick in the original ratings. It is Kumasaka's review rating that we shall report. There was also a series of training sessions in the Stirling rating system for three of our psychiatrists. These were conducted by Dr. Daniel O'Connell, a colleague of Dr. Alexander H. Leighton and Dr. Dorothea Leighton, who developed and made the psychiatric ratings in the Stirling County Study. O'Connell had been trained by them to be reliable with them in the ratings. The training materials were 48 interview protocols from the Stirling County research. All three of our psychiatrists achieved reliability with O'Connell on these protocols that was slightly higher than that reported above in footnote 31 for the tape spot check of interviewer "caseness" ratings. At a final training session, one of our psychiatrists, Dr. Gladys Egri, and O'Connell independently rated a sample of 22 interview protocols from the Washington Heights study. Their agreement was found to be remarkably high—72.7 per cent complete agreement and one step disagreement on all the rest. It is a consensus rating of O'Connell and Egri after each made independent review ratings that we shall report.

33. These consensus ratings were made by Dr. Gladys Egri and Dr. Daniel O'Connell after each had rated the protocols independently (see reference 32 above).
34. These ratings were made by Dr. Yorihiro Kumasaka (see reference 32 above).
35. These statistical tests of the reviewer ratings, unlike the tests for the interviewer ratings, are one-tailed because they test whether the original relationships obtained with the interviewer ratings hold.
36. These ratings were made by Dr. Daniel O'Connell (see reference 32 above).
37. Leighton, Dorothea C., et al. *Op. cit.*, p. 128.
38. See, for example, Srole, Leo, et al. *Op. cit.*, pp. 149, 333-334; Leighton, Dorothea C., et al. *Op. cit.*, pp. 128, 143.

Dr. Dohrenwend is Associate Professor, Department of Psychiatry, Columbia University Social Psychiatry Research Unit (128 Fort Washington Ave.), New York, N. Y. 10032

This work was supported by Grant MH 10328 from the National Institute of Mental Health, U. S. Public Health Service.

This paper was presented before the Mental Health Section of the American Public Health Association at the Ninety-Seventh Annual Meeting in Philadelphia, Pa., November 10, 1969.